International Conference on Signals and Electronic Systems
September 13-15, 2004 Poznań, POLAND
ICSES'04

An Area-Efficient FPGA Implementation of a Disparity Estimation Scheme for Real-Time Compression of IP Images*

D. Chaikalis[#], D. Maroulis[#], N. Sgouros[#], P. Papageorgas[##], N. Theofanous[#]

[#]National and Kapodistrian University of Athens, Dept. of Informatics and Telecommunications,
Panepistimiopolis, Ilisia, Athens 15784, GREECE
e-mail: d.xaikalis@di.uoa.gr, dmarou@di.uoa.gr, nsg@di.uoa.gr, optel@di.uoa.gr

[##]Technological Educational Institute of Piraeus, Dept. of Electronics,
Petrou Ralli & Thivon ave., Egaleo, Athens 12244, GREECE
e-mail: ppapag@teipir.gr

Abstract – **The interest for real three dimensional displays has been revived over the past few years, and the high resolution images that capturing and reproducing of 3D imagery demands, must be addressed with an efficient compression scheme. An area-efficient hardware implementation of a real-time disparity estimation scheme targeted to Integral Photography 3D images is presented, exploiting the inherent redundancy of Integral Photography images, coping with the increased bandwidth that this technology requires.**

Keywords: three dimensional display, autostereoscopic display, integral photography, hardware implementation, real-time, field programmable gate array,

I. INTRODUCTION

Today, two-dimensional colour imagery is the visual medium used in most applications. However, there exist many applications, both in person-to-person and in broadcast-type communication, which would greatly benefit from an increased degree of realism. The perception of depth, so natural in daily life, would greatly enhance a "being there" experience [1].

Over the past few years, the rapid increase in processing power, the widespread use of high resolution TFT-LCD displays, combined with improvements in manufacturing of high quality microoptics, revived the interest for three dimensional (3D) applications. Many promising technologies evolved, ranging from polarizing glasses, to most sophisticated techniques like shuttering glasses [2] and more recently autostereoscopic display devices [3].

Autostereoscopic display devices provide 3D stereoscopic viewing without the need of additional eyewear, while most of them allow multiple viewers to experience the 3D effect. A special category of autostereoscopic displays is based on the principles of Integral Photography (IP), first introduced by Lippman [4] back in 1908. One of the most important advantages of this technique is the natural viewing of 3D scenes without the common side effect of eye-strain. Such autostereoscopic displays can be of great value in medical [5], educational and entertainment [6] applications.

One of the main concerns in developing such applications is the necessity to cope with high resolution images that result in high bandwidth and storage requirements for the capturing and reproduction of 3D objects and scenes. Consequently a high-efficiency compression scheme of the associated data is crucial.

In Fig. 1 a number of sub-images are presented, where each sub-image corresponds to one microlens created with the image information of the underlying group of pixels of a high resolution TFT display. The information redundancy in the neighbouring sub-images is evident.

* This work was realized under the framework of the operational Program for Education and Vocational Training Project "Pythagoras", co-funded by European Union and the Ministry of National Education of Greece.

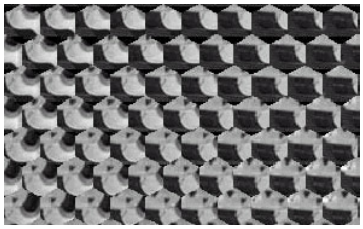

Fig. 1: The inherent redundancy present in an IP image

Hardware implementations can exploit pipelined and massively parallel processing, thus being an efficient solution for accelerating time-critical tasks, such as motion estimation and compression schemes, especially for real-time applications. Although Field Programmable Gate Arrays (FPGAs) and Application Specific Integrated Circuits (ASICs) have both been used for hardware implementation, the use of FPGAs has dominated over ASICs in the research and development phase the last years [7]. The primary benefits that the FPGAs offer are the following [8]:

- Increased flexibility: the functionality of the embedded processor can be quickly changed and design faults can be quickly rectified.
- Sufficient performance: the performance of FPGAs has increased tremendously and is quickly approaching that of ASICs.
- Faster design time: faster design times are achieved by re-using intellectual property cores and utilization of high-level hardware description languages (e.g., VHDL).

In this paper, a hardware implementation of a real-time disparity estimation scheme for IP images is presented, targeted to real-time 3D display and capturing applications. The key point is to take advantage of the inherent redundancy that is present in IP images, and cope with the bandwidth limitations of such an implementation, creating an area-efficient and bandwidth-limited hardware system.

II. MOTION ESTIMATION

One of the critical parts of the compression scheme followed is motion estimation. Motion estimation has been introduced in an attempt to trace the motion of objects within a video scene, i.e. find the best match between the pixels in the current frame and the pixels in the reference frame. To this end, a search area within the reference frame must be traversed in order to find the best match. After that, the intensity differences between the pixels must be coded together with the difference in coordinates between the locations (motion vector). Block-based motion estimation is the most commonly used form, in which a search is performed in the reference frame for a block of pixels in the current frame.

Two key issues are associated with motion estimation in general, namely the size of the search area and the metric to use for determining the "best match". For both issues, many methods have been proposed [9] in order to reduce the number of computations.

The most commonly used metric is the Sum of Absolute Differences (SAD), which adds up the absolute differences between corresponding elements in the macroblocks. The SAD value calculation is time-consuming due to the complex nature of the absolute operation and the subsequent multitude of additions. For a 8x8 block, the SAD is calculated as follows:

$$SAD(x,y,r,s) = \sum_{i=0}^{7}\sum_{j=0}^{7}\left|\left(A_{(x+i,y+j)} - B_{((x+r)+i,(y+s)+j)}\right)\right|$$

with

$0 \le x, y < framesize$

(r,s) being the motion vector

$A_{(x,y)}$ being a target frame pixel at (x,y)

$B_{(x,y)}$ being a reference frame pixel at (x,y)

III. COMPRESSION METHOD

The proposed hardware implementation is based on the compression method described in [10]. In this section, the outline of this method is presented.

The compression is applied on an IP image rather than an image sequence, which is the normal procedure followed for video compression. The goal is to exploit the correlation properties of the IP images, since neighbouring windows exhibit high spatial information redundance. Hence the term disparity estimation will be used hereafter instead of motion estimation.

The size of the search area is defined by the properties of the IP images. It is shown that a unidirectional search is capable of providing a best match very close to most standard block search methods, when used for IP images. An exhaustive block search method is used, targeted to high quality image requirements.

An intra sub-image (I) of size 32x32 is encoded in a JPEG like manner by using a DCT transform and proper quantization, followed by efficient coding to optimize performance. More specifically, an I sub-image is segmented to 16 8x8 pixel blocks and a 2D DCT is applied on each block using the quantization and coding strategy as described in the JPEG standard [11]. The sizes of these windows are realistic for a normal IP imaging device.

Based on the encoded I sub-image, an estimation of two predicted sub-images (P) of 32x32 pixel size is formed by estimating the proper disparity vectors for each P sub-image.

In order to calculate the disparity vectors, the SAD metric is used for determining the best match of each 8x8 pixel block of each P sub-image using an I sub-image as reference. SAD calculation and block matching procedures are implemented into hardware, in order to

achieve real-time estimation. In Fig. 2, the sub-image placement in the image and the block search method are depicted.

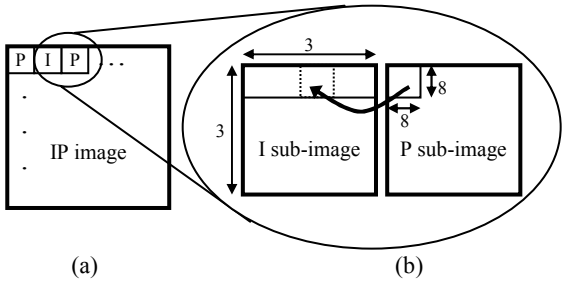

Fig. 2: (a) The placement of P and I sub-images in an IP image, (b) block search area outline

IV. HARDWARE DESIGN

For the proposed hardware design, VHDL (Very High Scale Integrated Circuits (VHSIC) Hardware Description Language) is used. The design was verified and simulated using the Xilinx ISE 5 development software, along with the ModelSim simulator. The hardware modules were implemented in FPGA using the Celoxica RC1000-PP PCI board.

The FPGA on the development board is a Xilinx Virtex XCV-2000E chip, with an area equivalent of 2 million gates. The board includes 8 MB memory modules, and the FPGA implements a dedicated 640kbit dual-ported memory, which can be used in any width desired [12]. The development board has the ability to host entire images in its memories. Downloading entire images to the board speeds up calculations, taking in concern that transfers to and from the PCI board slow down overall performance and should be as fewer as possible.

The SAD calculations can be performed in a completely parallel manner, increasing the overall speed of the system. Such an approach though poses the problem of high bandwidth demands, for the solution of which multiple FPGAs should be used [13]. In the proposed method, we implement only 8 SAD units for each block comparison, in order to create an area-efficient hardware implementation for the specific FPGA. The hardware design of the SAD units and the adder tree are based on [9], where a SAD implementation in FPGA hardware is presented. For the absolute difference calculation, the hardware system inverts the smallest value of each pixel pair, then adds 8 pairs with a 17-to-1 full-adder adder tree and a correction term is added in the end, so as to have the correct value as an output.

For the exhaustive unidirectional search to be performed, 51 comparisons are needed for each line of blocks in the sub-image. For example, for the leftmost block, 25 8x1 SAD units are operating in parallel, receiving the pixel values as the P and I sub-image lines are scanned from left to right in a group of 8. The 24 of these units need a delay register before each of their inputs, so that the data are successfully aligned. In the same manner, when the scanning of each line reaches the 8th column, 17 8x1 SAD units operate on behalf of the second 8x8 block, and when the scanning reaches the 15th column, 7 units are used for the third 8x8 block. In the current framework, no comparison is performed for the rightmost block. Therefore, by using 25+17+9 8x1 SAD units, we are able to calculate the minimum SAD value for the 4 8x8 blocks of each line of the P sub-image.

As a result, 4 blocks of the P sub-image are estimated by a single scanning of 8 lines of the P and I sub-images. For each block, a comparison unit is needed so as to determine the minimum SAD value and the disparity vector. The comparison unit compares its input value with the previous value stored in its register, and stores the smaller of the two. The comparison between two values needs only one clock cycle, and it can be performed simultaneously with the SAD calculations, since each 8x1 SAD unit produces a result one clock cycle after the previous one. For each comparison, the comparison unit also stores the position offset of the I sub-image's block with the minimum SAD value, namely the disparity vector.

After the appropriate number of iterations, depending on image size, the FPGA system completes its operations and sends the generated data to the host computer. The data consists of the disparity vector matrix for each P sub-image, and the 8x8 blocks with the intensity differences between the I and P sub-image blocks for which the SAD value is the minimum. The schematics for the 8x1 SAD unit and the 8x8 block comparison units are illustrated in Fig. 3.

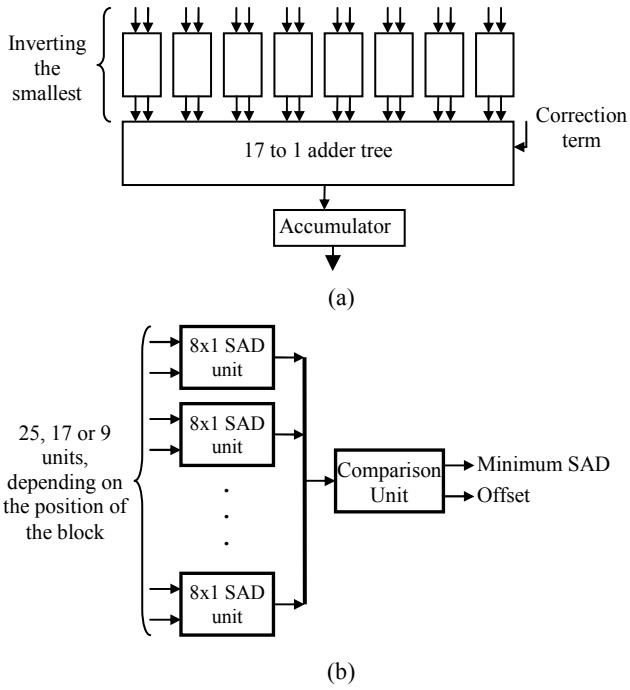

Fig. 3: (a) the 8x1 SAD unit, (b) the 8x8 block comparison unit

V. RESULTS

In order to create an area-efficient system, which is our main target, we focused our hardware implementation to

a single FPGA, taking into consideration the bandwidth limitations of such an approach. Using 51 8x1 SAD units as described above, in order to simultaneously calculate the minimum SAD for 4 8x8 blocks, along with the appropriate delay registers and the 3 comparison units needed, approximately 65,5% of the Xilinx Virtex XCV-2000E FPGA area is occupied. This hardware system implementation needs two 1kbyte 64-bit wide memories, which are implemented using the dedicated dual-port memories that the specific FPGA offers.

Our hardware implemented system can be clocked with a maximum frequency of 60.6 MHz. This is justified by the lack of pipeline stages in the 8x1 SAD module, in order to drastically decrease the area demands. For this clock rate, it is capable of calculating the minimum SAD for 4 blocks in 34 clock cycles, thus needing 136 clock cycles for each P sub-image. This results in 445632 P sub-images that can be processed every second.

In Fig. 4, a number of standard image sizes and the relevant number of IP images that can be processed is presented.

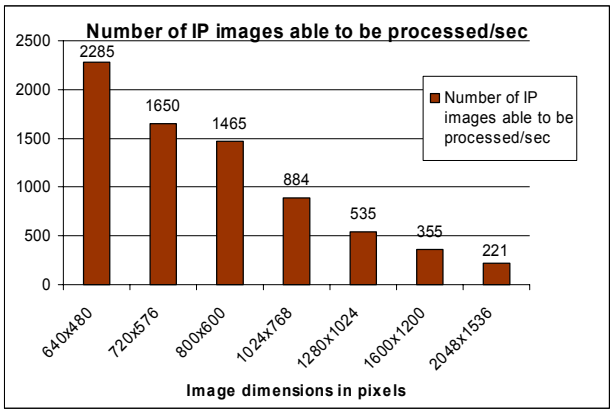

Fig. 4: Number of IP images able to be processed, according to image dimensions

It is obvious that the hardware implementation can process in real-time IP images of significant size. This is very useful since IP deals with high quality images. It should be noted though that additional time is needed for transfers to and from local memories, and also for FPGA-to-PC transfers, which require a significant time quota, since the bandwidth for these transfers is limited by the PCI bandwidth of the FPGA board utilized.

VI. Conclusions – Future work

In this paper, the problem of efficiently generating estimation disparity vectors for IP images for use in real-time 3D display and capturing systems is addressed. A software approach is analyzed and a hardware implementation of a block comparison unit, based on the SAD metric, is presented. The system designed sacrifices some speed performance in order to achieve the bandwidth demands which allow implementation into a single FPGA, in contrast with other SAD implementations.

Other components, such as a compression module and coders that are time-consuming when realized in software, will be evaluated for efficient hardware implementation, in order to create a complete compression scheme specifically targeted to IP images. It is of importance to mention that the same compression scheme and hardware implementation can be used in most of the technologies used in autostereoscopic displays resulting to a universal methodology for compression of the corresponding images.

References

[1] Janusz Konrad "Visual Communications of Tomorrow: Natural, Efficient and Flexible", *IEEE Communications Magazine,* vol. 39, no. 1, pp. 126-133, Jan. 2001

[2] *Developers Handbook,* Stereographics 1997, www.stereographics.com

[3] M. Halle, "Autostereoscopic Displays and Computer Graphics", *Computer Graphics,* 31(2), ACM SIGGRAPH, pp. 58-62, 1997

[4] G. Lippman, "La Photographie Integrale", *C. R. Acad. Sci,* 146, pp. 446-451, 1908

[5] H. Liao, S. Nakajima et. al., "Intra-operative Real-Time 3D Information Display System Based on Integral Videography", *MICCAI' 01,* LNCS 2208, pp. 392-400, 2001

[6] P. Harman, "Home Based 3D entertainment – an overview", *Proc. ICIP(1)*, pp. 1-4, 2000

[7] S. Ramachandran, S. Srinivasan, "FPGA Implementation of a Novel, Fast Motion Estimation Algorithm for Real-Time Video Compression", *9th Int. Symposium on FPGAs,* Monterey, Canada, 2/2001

[8] S. Rathnam, G. Slavenburg, "An Architectural Overview of the Programmable Multimedia Processor, TM-1", *Proc. COMPCON '96,* pp. 319-326, 1996

[9] S. Wong, S. Vassiliadis, S. Cotofana, "A Sum of Absolute Differences Implementation in FPGA Hardware", *28th Euromicro Conference (EUROMICRO'02),* pp. 183-186, Dortmund, Germany, September 04 - 06, 2002

[10] N. Sgouros, A. Andreou, M. Sangriotis, P. Papageorgas, D. Maroulis, N. Theofanous, "Compression of IP Images for Autostereoscopic 3D Imaging Applications", *3rd International Symposium on Image and Signal Processing and Analysis (ISPA03),* Rome, Italy, September 18-20, 2003

[11] W. B. Pennebaker, J. L. Mitchell, *JPEG Image Compression standard,* VNR, NY, 1993

[12] Celoxica RC1000-PP development board: Hardware Reference, www.celoxica.com

[13] S. Wong, B. Stougie, S. Cotofana, "Alternatives in FPGA-based SAD Implementations", *IEEE I.C. on Field Programmable Technology 2002 (FPT '02),* Hong Kong, December 2002